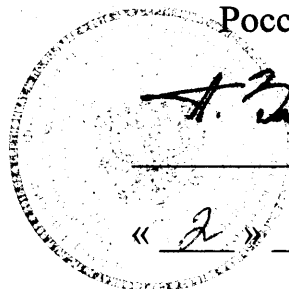
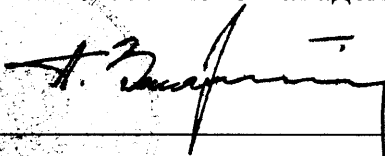


УТВЕРЖДАЮ
Директор Федерального
государственного учреждения
«Федеральный исследовательский
центр «Информатика и управление»
Российской академии наук»




И.А.Соколов
« 2 » августа 2017 г.

ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

на диссертацию Хандеева Владимира Ильича «Алгоритмы с оценками качества для квадратичных задач кластеризации с фиксированным центром одного из кластеров», представленную на соискание ученой степени кандидата физико-математических наук по специальности 01.01.09 – Дискретная математика и математическая кибернетика

Анализ сложности различных труднорешаемых дискретных оптимизационных задач и методов их решения является в настоящее время весьма значимым для развития производственной, социальной и научной сфер нашей жизни. Особую значимость имеет оценка качества алгоритмов минимизации квадратичных критериев кластеризации. Рассматриваемые в диссертации оптимизационные задачи близки к NP-трудной в сильном смысле известной задаче MSSC (Minimum Sum of Squares Clustering), в которой требуется разбить конечное множество векторов на несколько кластеров с минимальной суммой внутрикластерных сумм квадратов расстояний до их центроидов. В работе Хандеева В.И. рассматриваются квадратичные задачи кластеризации со специальным ограничением (для одного из кластеров рассматривается сумма квадратов расстояний от его элементов до начала координат), в том числе задачи кластеризации последовательности. На данные оптимизационные задачи кластеризации множества и последовательности точек непосредственно не переносятся результаты для задачи MSSC. Здесь требуется создание новых математических методов. Тема диссертации весьма актуальна и в связи с большой, широкой областью практических применений, развитием Интернета, созданием хранилищ данных. Возникли задачи «больших данных (big data)» и «разработки данных (data mining)», которые в принципе не поддаются решению на современных компьютерах. Таким образом, данное исследование сложности решения распространенных задач дискретной оптимизации и изучения вопросов их аппроксимируемости является весьма **актуальным**, а

создание полиномиальных и псевдополиномиальных схем их решения является необходимым в нашей жизни.

Диссертация В.И. Хандеева состоит из введения, трех глав и заключения.

Во **введении** показана актуальность работы, ее важность для многих научных и технических приложений, цель работы (создание эффективных алгоритмов с гарантированными оценками качества для одной модификации MSSC и кластеризации последовательностей), приводятся основные результаты, отмечены научная новизна, личный вклад автора, теоретическая значимость, практическая ценность диссертации, приведен перечень конференций и семинаров, где докладывались результаты.

В **первой** главе изложены предпосылки рассматриваемых задач, приводятся полученные ранее наиболее важные математические результаты по общей задаче MSSC, в том числе приближенная полиномиальная схема PTAS для случая фиксированного числа кластеров и размерности описаний точек. Следует отметить хорошее знание диссертантом современного уровня результатов в данной области. Приводится много ссылок на литературу и разных оценок. Следует отметить, что Хандеев В.И. приводит обзор результатов не только в первой главе, но и критически рассматривает последние результаты, относящиеся к рассматриваемым задачам в главах 2 и 3. Здесь же обоснована задача разбиения конечного множества объектов на два множества (соответствующих активному и пассивному состояниям объекта). Отмечена слабая изученность оптимизационных задач кластеризации. Рассмотрены публикации по кластеризации упорядоченных по времени данных. Данная задача также имеет многочисленные приложения (в частности, помехоустойчивый анализ временных рядов). Ее формализация по критерию минимума суммы квадратов отклонений индуцирует рассматриваемую в главе 3 оптимизационную дискретную задачу 2-разбиения, отличающуюся от задач главы 2. Удачны и рисунки 1.1-1.6, иллюстрирующие рассматриваемые задачи 2-MSSC.

Во **второй** главе изложены полученные результаты по квадратичной задаче 2-кластеризации конечного множества векторов при фиксированном центре одного из кластеров. При этом исследуются два варианта: с оптимизируемыми мощностями кластеров и с ограничениями на мощности кластеров. Для первой задачи создан 2-приближенный полиномиальный алгоритм A_1 (пункт 2.1.3.) и доказана теорема 2.1. Для второй задачи получены точный алгоритм для специального случая, аппроксимационная схема и рандомизированный алгоритм. Доказано, что точный алгоритм при фиксированной размерности пространства является псевдополиномиальным. Показано, что для задачи не существует схемы FPTAS (теорема 2.3), и такая схема (алгоритм A_3) предложена для

фиксированной размерности пространства. Здесь же создан рандомизированный алгоритм, который находит (при заданной относительной ошибке и вероятности несрабатывания) приближенное решение за линейное время от размерности пространства и от числа входных точек. Получены условия, когда алгоритм асимптотически точен и имеет небольшую трудоемкость. Отметим, что созданные алгоритмы и доказательство ряда теорем (особенно теоремы 2.5) не просты и опираются на ряд вспомогательных лемм.

В **третьей** главе содержатся результаты исследований сложности квадратичных задач кластеризации конечных последовательностей точек при фиксированном центре одного кластера, при этом используются результаты второй главы. Получены результаты для задачи 2-кластеризации, а также для задач многокластерного разбиения с ограничениями и без ограничений на мощности искомых кластеров. Здесь широко применяется динамическое программирование. Предложен точный алгоритм для задачи 2-кластеризации при целочисленных входах. Показана псевдополиномиальность алгоритма при ограниченности размерности пространства. Здесь же установлено (при известной гипотезе $P \neq NP$), что для общего случая не существует схемы FPTAS, создан приближенный алгоритм, доказана при фиксации размерности пространства реализация с его помощью схемы FPTAS, обоснован рандомизированный алгоритм. Алгоритмы 2-кластеризации последовательности имеют те же (или полиномиально увеличенные) вероятностные и аппроксимационные характеристики, что и алгоритмы 2-кластеризации множеств точек. Для квадратичной евклидовой задачи многокластерного разбиения конечной последовательности с ограничениями на выбор внутрикластерных точек созданы 2-приближенные алгоритмы для задач без ограничений и с ограничениями на мощности кластеров, показана их полиномиальность для фиксированного набора кластеров. Таким образом, в главе 3 рассмотрен целый комплекс задач, индуцированных основной проблемой, предложены алгоритмы и получены оценки их качества.

В **заключении** сформулированы основные результаты, полученные в диссертационной работе. Отмечено, что перспективным направлением дальнейших исследований является создание более быстрых алгоритмов с оценками качества, в том числе линейных и сублинейных.

Таким образом, в диссертации получены следующие **основные результаты**.

1. Для квадратичной задачи кластеризации конечного множества точек евклидова пространства на два кластера с фиксированным центром одного кластера для задачи без ограничений на мощности кластеров построен 2-приближенный полиномиальный алгоритм, а при ограничениях предложен рандомизированный алгоритм, находящий

приближенное решение за линейное время. Найдены условия его полиномиальности и асимптотической точности.

2. Для евклидовых квадратичных задач 2-кластеризации конечных множеств и последовательностей (с ограничениями на выбор элементов) при фиксированном центре одного кластера и ограничении на мощность кластеров построены точные алгоритмы для целочисленных входов задачи, псевдополиномиальные при фиксированной размерности пространства.
3. Показано, что для общих случаев задач из п. 2 не существует схем FPTAS (если $P \neq NP$) и такие схемы построены для задач с фиксированной размерностью пространства.
4. Для многоклассовой квадратичной задачи разбиения последовательности точек с ограничениями на их выбор и при фиксированном центре одного кластера созданы 2-приближенные алгоритмы для задач с ограничениями и без ограничений на мощности классов. Показано, что алгоритмы полиномиальны при заданном числе кластеров.

В целом диссертация выполнена на высоком научно-техническом уровне, написана хорошим языком и оформлена на уровне современных редакторских возможностей. Принципиальные недостатки в работе отсутствуют. Тем не менее, по диссертации имеются следующие **замечания и пожелания**:

1. На наш взгляд, в обзоре известных результатов по задаче 1 при упоминании задачи 3-SAT было бы целесообразно привести хотя бы расшифровку её названия; здесь же уместно было бы дать определения аппроксимационных схем PTAS, FPTAS, а также рандомизированных и асимптотически точных алгоритмов.
2. Доказательство леммы 2.10 содержит корректные, но не совсем очевидные переходы – для удобства читателя их можно было бы снабдить более подробными комментариями.
3. Многомерная кубическая решетка в пункте 2.2.3 – частный случай решетки, определенной в следующем пункте 2.2.4. Поэтому можно было бы сразу ввести общее определение, чтобы избежать дублирования. Это же общее определение может быть использовано при формулировке точного псевдополиномиального алгоритма в третьей главе.

Вместе с этим, отмеченные замечания не снижают научно-практической ценности и значимости выполненной Хандеевым В.И. диссертационной работы.

Автореферат и публикации автора **полностью** отражают сформулированные в диссертации научные положения и выводы. Диссертация посвящена исследованию новых задач, многие результаты получены автором впервые. Например, алгоритмы из пунктов 1 и 2

основных результатов либо имеют меньшую трудоемкость по сравнению с известными, либо реализуют новые перспективные подходы к решению задач. Впервые построены некоторые оригинальные аппроксимационные схемы, установлен факт несуществования схемы FPTAS для рассматриваемых задач. Для специальных задач кластеризации последовательностей построены единственные на сегодняшний день приближенные алгоритмы с гарантированными оценками точности.

Результаты диссертации опубликованы в 9 работах из списка ВАК, доложены на многочисленных международных и всероссийских конференциях, семинарах ИМ СО РАН и НГУ. Научные положения, выносимые на защиту, **полностью** отражены в опубликованных работах. Язык и стиль диссертации и автореферата соответствуют требованиям ВАК, участие соавторов в совместных публикациях отмечено точно и в нужной мере, соответствие поставленных целей и полученных результатов полностью выполнено.

Считаем, что полученные в диссертационной работе результаты представляют теоретический и практический интерес для научно-исследовательских, учебных организаций (ФИЦ ИУ РАН, ИМ СО РАН, ИММ УрО РАН, ИПМ РАН, ИСП РАН, МГУ, НГУ, МФТИ, и др.) и отдельных исследователей, занимающихся оценками сложности задач и алгоритмов дискретной оптимизации, созданием приближенных эффективных схем. Полученные результаты могут быть использованы в данных организациях и в учебном процессе в Университетах РФ, осуществляющих подготовку специалистов в области дискретной математики. Считаем целесообразным продолжить работу по исследованию сложности задач дискретной оптимизации, созданию полиномиальных приближенных алгоритмов их решения.


На основании ознакомления с содержанием диссертации, автореферата, опубликованных автором работ можно сделать следующее **заключение**: диссертационная работа Хандеева Владимира Ильича выполнена на актуальную тему, полученные результаты – значимы и обоснованы доказанными леммами и теоремами, она является законченным научным исследованием. Работа отвечает критериям, установленным Положением о порядке присуждения ученых степеней, утвержденном Постановлениями Правительства Российской Федерации от 30.01.2002 г. № 74 (в редакции постановления Правительства Российской Федерации от 20.06.2011 г. № 475, а также в редакции Постановления Правительства РФ от 21.04.2016 № 335). Проведенные научные исследования можно характеризовать как научно-обоснованные математические разработки, обеспечивающие решение важных теоретических задач в области создания эффективных алгоритмов решения дискретных оптимизационных задач кластеризации.

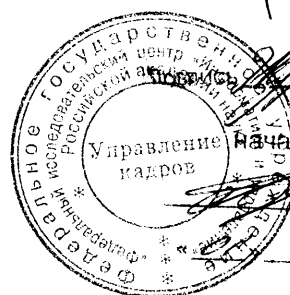
Это дает основание полагать, что Хандеев В.И. заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 01.01.09 – дискретная математика и математическая кибернетика.

Отзыв на диссертацию и автореферат обсуждены на заседании Отделов «Распознавание, защита и анализ информации» и «Методы классификации и анализа данных» Вычислительного центра им. А.А.Дородницына ФИЦ ИУ РАН (протокол № 01-08-2017/01 от 01 августа 2017 года).

Отзыв составлен Рязановым Владимиром Васильевичем, зав. отделом «Методы классификации и анализ данных», тел. 8 (916) 912 8820, e-mail: rvccas@mail.ru

Заместитель директора Вычислительного центра им. А.А. Дородницына ФИЦ ИУ РАН, заведующий отделом «Распознавание, защита и анализ информации» академик РАН

 Ю.И. Журавлев

 заверяю:
Начальник отдела
А.Н. Кудряшова
08 2017